

Generating a Synthetic Population in Support of Agent-Based Modeling of Transportation in Sydney

N. Huynh^a, M. Namazi-Rad^a, P. Perez^a, M. J. Berryman^a, Q. Chen^a, and J. Barthelemy^a

^a SMART Infrastructure Facility, University of Wollongong, NSW 2522, Australia

Email: mohammad_namazi@uow.edu.au

Abstract: The complexity of large cities such as Sydney makes planning challenging. There is a growing need for new and evolving tools to assist research and decision-making. Increasingly, planners require sophisticated insights on social behaviour and the interdependencies characterising urban systems. Agent-based modelling as a large and wide-spread scientific modelling technique (that focuses on computer modelling of individuals and their interactions) has recently emerged as a promising tool in this regard with applications to real-world problems in infrastructure, particularly transport planning, of urban areas. An essential element of such an agent based model is a realistic synthetic population that matches the distribution of individuals and households living in a study area as per the demographics from census data. This paper presents an algorithm to construct such a synthetic population that uses only aggregated data of demographic distributions as inputs, and an agent based model which simulates the natural evolutions (ageing, marriage, divorce, reproducing) of this initial population. The significance of the synthetic population developed in this work is in its ability to capture the relationship of individuals in a household and changes in structure of households as individuals undergo natural evolutions. A case study that uses the algorithm to initialise a synthetic population for Randwick (Sydney) in 2006 and evolve this population over 5 years will also be presented. The results of the initial and final population were validated against the Census Data in 2006 and 2011. The paper closes with discussions on the application of this synthetic population to simulate the dynamics interaction between transport and land-use.

Keywords: *Agent-Based Modelling, Combinatorial Optimization Model, Hierarchical Structure, Household Dynamics; Population Synthetiser.*

1. INTRODUCTION

The increasing population in Australia is highly urbanised and the complexity of large cities such as Sydney makes planning challenging. Increasingly, planners require sophisticated insights on social behaviour and the interdependencies characterising urban systems. Agent-based modelling as a large and wide-spread scientific modelling technique (that focuses on computer modelling of individuals and their interactions) has recently emerged as a promising tool in this regard with applications to real-world problems in infrastructure, particularly transport planning, of urban areas. An essential element of such an agent based model is a realistic synthetic population that matches the distribution of individuals and households living in a study area as per the demographics from census data.

A large number of works on the construction of synthetic population microdata for small area scale has been adapting one of the two major approaches, the synthetic reconstruction (SR) and the combinatorial optimisation (CO). The evaluation and comparison of these two approaches, as well as detailed descriptions of each of them, have been in several publications. Examples include Huang and Williamson (2001), Ryan *et al.* (2007), Muller and Axhausen (2011a), and Kurban *et al.* (2011).

The SR approach was first presented by Wilson and Pownall (1976), and has since appeared to be the more popular approach. The approach normally comprises a sequential process of random sampling from a set of conditional probabilities, which is usually derived from cross tabulations of the desired attributes of the population [Huang and Williamson, 2001]. These cross tabulations of demographic attributes of the population are sourced from census data, which is often the only available source of population data at small area scale (e.g. at census district or travel zone level and comprising few hundred households). Yet it is highly unlikely that the cross tabulations provide a complete description of the interdependency of all the desired synthetic population attributes. For example, if the synthetic population needs four demographic attributes a , b , c , d , published crosstabs in the census data may provide only partial joint probability distribution of these attributes, say $Q1(a,b,c)$, $Q2(b,d)$, and $Q3(a,d)$. To overcome this issue, the well known iterative proportion fitting (IPF) (Deming and Stephan, 1940) or its variations (Gou and Bhat, 2007; Ye *et al.*, 2009; Muller and Axhausen, 2011b) are commonly used to link data contained in separate tables, and thus provide a full joint probability distribution for four variables a , b , c , and d while still maintaining the constraints $Q1$, $Q2$, and $Q3$. If a desired attribute is missing from the census data, the missing data can be derived from other sources and incorporated into the existent data to form a full joint probability distribution as described above. Such ability to incorporate data from various sources has been reported as one of the advantages of the use of conditional probabilities in the synthetic reconstruction approach. [Huang and Williamson, 2001]. Once the full joint probability distribution of the desired population attributes are known, the synthetic population of households can be generated by drawing entire households from a source of disaggregate data (which is normally from a survey over a sample of the population) following the estimated probabilities given in the full joint probability distribution. Variations of the approach exist, which mainly differ in the way the iterative proportion fitting procedure is implemented (Beckman *et al.*, 1996; Frick and Axhausen, 2004; Gou and Bhat, 2007; Ye *et al.*, 2009, Auld *et al.*, 2009, Muller and Axhausen, 2011b; Huang and Williamson, 2001, Barthelemy and Toint, 2012).

An alternative to the SR method is the combinatorial optimisation method, first presented by Williamson (1998). The process first randomly picks a group of individuals (or a group of households) from a disaggregate data set (e.g. a survey over a sample of the population over a large region) so that it matches the population size of the small area. The observed sample is then statistically compared against a pre-defined set of demographic characteristics of the small area. An example of the statistic used in such a comparison is the modified overall Relative Sum of Squared Z-scores (RSSZm) (Huang and Williamson, 2001). If the goodness of fit is not satisfactory, a record (either a household or an individual) in the observed sample is swapped with one from the pool (i.e. the survey). The statistical comparison is then carried out for the new observed sample. This process is repeated with the aim being gradually improving the goodness of fit and is stopped when a critical value of the comparison statistics is met. Given the search space (the survey data), the final synthetic population from this approach is normally the best achievable at a given time, rather than a guaranteed optimal solution (Huang and Williamson, 2001). The first synthetic population generators adapting this approach were done by the group led by Williamson and co-workers (Williamson *et al.*, 1998; Voas and Williamson, 2000; Huang and Williamson, 2001). Another major research effort to build a synthetic population using the approach has been carried out at National Centre for Social and Economic Modelling (NATSEM) based at the University of Canberra, Australia (Harding *et al.*, 2004; Williams, 2003; Melhuish *et al.*, 2002).

One issue that has not been addressed in all aforementioned population synthesisers is the constraints of the composition of residents in households by their relationships. Such a synthetic population, in which

individuals are grouped into households by their relationship (and make decisions collectively), is critical in an agent based model that captures realistic behaviours of individuals in their interactions with infrastructure systems. For example a household with a single parent with two children under 15 years old would have a considerably different transport need and behaviours than a married couple household with no kids. More importantly, such a synthetic population (in which households are structured by relationship of residents) allows the simulation of natural evolution of the population, which helps provide a better projection of the demographics of future population.

This paper presents a methodology to initialise and evolve a synthetic population in which the individuals are simulated discretely and are grouped into households based on constraints of the composition of residents in households by their relationships (Section 2). The algorithm to initialise such a synthetic population does not require a disaggregate data set of household or individual records and uses only aggregated data of demographic distributions as inputs. These distributions are available in the census data published by the Australian Bureau of Statistics (ABS). The evolution of this population captures the ageing, coupling, divorcing, and reproducing of individuals as well as the consequent changes in structure of households as a result of these processes. A case study that uses the algorithm to initialise a synthetic population for Randwick area in Sydney in 2006 and evolves this population over 5 years will also be presented. The results of the initial and final population were validated against the ABS census data in 2006 and 2011. The paper closes with discussions on the application of this synthetic population to simulate the dynamics interaction between transport and land-use (Section 3).

2. THE POPULATION SYNTHESISER

2.1 Synthetic population construction algorithm

The algorithm presented in this work to construct a synthetic population for agent based modelling purposes is a variation of the CO approach as described in the previous section. Individuals are selected from an individual pool and allocated into households in a household pool to satisfy the distribution of household compositions in a study area. Different to traditional CO approach, the pool of individuals (i.e. a set of disaggregate records of individuals) is instantiated from an aggregate data set representing the demographics distribution of the study area rather than extracted from an existent disaggregate survey data. The pool of households is instantiated from another aggregate data set. The construction of the individual pool and the household pool, as well as the allocation of individuals into households will be presented in this section.

2.1.1 Constructing the individual pool and household pool

| | 0-14 years | 15-24 years | 25-34 years | 35-44 years | 45-54 years | 55-64 years | 65-74 years | 75-84 years | 85 years and over | Total |
|---|---------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------------|------------|
| MALES | | | | | | | | | | |
| Husband in a registered marriage | .. | 0 | 12 | 31 | 9 | 11 | 3 | 0 | 0 | 66 |
| Partner in de facto marriage(b) | .. | 3 | 31 | 14 | 11 | 4 | 3 | 0 | 0 | 66 |
| Lone parent | .. | 0 | 0 | 0 | 6 | 0 | 0 | 0 | 0 | 6 |
| Child under 15 | 33 | .. | .. | .. | .. | .. | .. | .. | .. | 33 |
| Dependent student (Aged 15-24 years) | .. | 5 | .. | .. | .. | .. | .. | .. | .. | 5 |
| Non-dependent child | .. | 0 | 3 | 0 | 6 | 0 | 0 | 0 | 0 | 9 |
| Other related individual | .. | 5 | 4 | 3 | 0 | 0 | 0 | 0 | 0 | 12 |
| Unrelated individual living in family household | .. | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 3 |
| Group household member | .. | 3 | 18 | 12 | 3 | 4 | 0 | 0 | 0 | 40 |
| Lone person | .. | 3 | 12 | 3 | 11 | 7 | 3 | 0 | 0 | 39 |
| Visitor (from within Australia)(c) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 33 | 19 | 80 | 66 | 46 | 26 | 9 | 0 | 0 | 279 |
| FEMALES | | | | | | | | | | |
| Wife in a registered marriage | .. | 0 | 25 | 25 | 7 | 10 | 0 | 0 | 0 | 67 |
| Partner in de facto marriage(b) | .. | 3 | 27 | 10 | 4 | 0 | 0 | 0 | 0 | 44 |
| Lone parent | .. | 0 | 0 | 3 | 3 | 0 | 3 | 5 | 0 | 14 |
| Child under 15 | 26 | .. | .. | .. | .. | .. | .. | .. | .. | 26 |
| Dependent student (Aged 15-24 years) | .. | 6 | .. | .. | .. | .. | .. | .. | .. | 6 |
| Non-dependent child | .. | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 |
| Other related individual | .. | 0 | 4 | 3 | 0 | 0 | 0 | 0 | 0 | 7 |
| Unrelated individual living in family household | .. | 4 | 3 | 3 | 0 | 0 | 0 | 0 | 0 | 10 |
| Group household member | .. | 7 | 16 | 3 | 3 | 5 | 0 | 0 | 0 | 34 |
| Lone person | .. | 0 | 14 | 7 | 8 | 12 | 4 | 0 | 0 | 45 |
| Visitor (from within Australia)(c) | 0 | 3 | 3 | 3 | 0 | 0 | 3 | 0 | 0 | 12 |
| Total | 26 | 29 | 92 | 57 | 25 | 27 | 10 | 5 | 0 | 271 |

Figure 1. Example of input table 'Relationship in Household by Age by Sex' from 2006 Census Community Profile Series for the construction of individual pool

These tables are available in the 2006 Census Community Profile Series. An example of these tables for a Census Collection District in Sydney is given in Figure 1 and Figure 2.

The aggregate datasets that were used to construct the pool of individuals were tables 'Relationship in Household by Age by Sex'. The aggregate datasets that were used to construct the pool of households were tables 'Family Composition', 'Family Composition by Sex of Person in Family', and 'Household Composition by Number of Persons Usually Resident'.

| Families | | | Males Females Persons | | | |
|---|------------|------|---|------------|------------|------------|
| Couple family with no children | 83 | HF1 | Couple family with no children | 94 | 76 | 170 HF1 |
| Couple family with: | | | Couple family with: | | | |
| children under 15 and: | | | children under 15 and: | | | |
| dependent students and non-dependent children | 0 | HF2 | dependent students and non-dependent children | 0 | 0 | 0 HF2 |
| dependent students and no non-dependent children | 3 | HF3 | dependent students and no non-dependent children | 4 | 5 | 9 HF3 |
| no dependent students and non-dependent children | 0 | HF4 | no dependent students and non-dependent children | 0 | 0 | 0 HF4 |
| no dependent students and no non-dependent children | 32 | HF5 | no dependent students and no non-dependent children | 54 | 53 | 107 HF5 |
| no children under 15 and: | | | no children under 15 and: | | | |
| dependent students and non-dependent children | 0 | HF6 | dependent students and non-dependent children | 3 | 0 | 3 HF6 |
| dependent students and no non-dependent children | 4 | HF7 | dependent students and no non-dependent children | 5 | 3 | 8 HF7 |
| no dependent students and non-dependent children | 0 | HF8 | no dependent students and non-dependent children | 0 | 4 | 4 HF8 |
| One parent family with: | | | One parent family with: | | | |
| children under 15 and: | | | children under 15 and: | | | |
| dependent students and non-dependent children | 0 | HF9 | dependent students and non-dependent children | 0 | 0 | 0 HF9 |
| dependent students and no non-dependent children | 0 | HF10 | dependent students and no non-dependent children | 0 | 0 | 0 HF10 |
| no dependent students and non-dependent children | 3 | HF11 | no dependent students and non-dependent children | 0 | 6 | 6 HF11 |
| no dependent students and no non-dependent children | 7 | HF12 | no dependent students and no non-dependent children | 8 | 6 | 14 HF12 |
| no children under 15 and: | | | no children under 15 and: | | | |
| dependent students and non-dependent children | 0 | HF13 | dependent students and non-dependent children | 0 | 0 | 0 HF13 |
| dependent students and no non-dependent children | 4 | HF14 | dependent students and no non-dependent children | 0 | 6 | 6 HF14 |
| no dependent students and non-dependent children | 6 | HF15 | no dependent students and non-dependent children | 11 | 6 | 17 HF15 |
| Other family | 10 | HF16 | Other family | 10 | 8 | 18 HF16 |
| Total | 152 | | Total | 189 | 173 | 362 |

(a) 'Family Composition'

(b) 'Family Composition by Sex of Person in Family'

| | Family households | Non-family households(c) | Total |
|-------------------------------------|-------------------|--------------------------|------------|
| Number of persons usually resident: | | | |
| One | .. | 77 | 77 |
| Two | 93 | 33 | 126 |
| Three | 36 | 9 | 45 |
| Four | 18 | 0 | 18 |
| Five | 4 | 0 | 4 |
| Six or more | 0 | 0 | 0 |
| Total | 151 | 119 | 270 |

(c) 'Household Composition by Number of Persons Usually Resident'

Figure 2. Example of input tables from 2006 Census Community Profile Series for the construction of household pool

In order to reserve the confidentiality of the census data, small random adjustments were introduced into these tables. As a result there are mismatches across the tables that characterise individual attributes and those characterising household attributes. For example, table 'Family Composition' in Figure 2 indicates there are no family households of type HF8 (a couple with no children under 15 and no students and at least 1 non-dependent child), while table 'Family Composition by Sex of Person in Family' shows that there are totally 4 persons living in this household type. The mismatches need to be corrected in these tables before they are used to construct individual pool and household pool. Assumptions that define the dependencies of attributes in tables in Figures 1 and 2 and are used as constraints throughout the correction are listed below.

Assumption 1 The number of parents and the minimum number of individuals of various household relationships in each household type is assumed as follows

Table 1. Minimum number of individuals in each household type by their relationships.

| | NF | HF1 | HF2 | HF3 | HF4 | HF5 | HF6 | HF7 | HF8 | HF9 | HF10 | HF11 | HF12 | HF13 | HF14 | HF15 | HF16 |
|---------------------|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|------|------|------|
| min_Parents | -1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | -1 |
| min_U15Child | -1 | -1 | 1 | 1 | 1 | 1 | -1 | -1 | -1 | 1 | 1 | 1 | 1 | -1 | -1 | -1 | -1 |
| min_Student | -1 | -1 | 1 | 1 | -1 | -1 | 1 | 1 | -1 | 1 | 1 | -1 | -1 | 1 | 1 | -1 | -1 |
| min_O15Child | -1 | -1 | 1 | -1 | 1 | -1 | 1 | -1 | 1 | 1 | -1 | 1 | -1 | 1 | -1 | 1 | -1 |
| min_Relative | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| min_NonFamilyMember | 1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| min_Residents | 1 | 2 | 5 | 4 | 4 | 3 | 4 | 3 | 3 | 4 | 3 | 3 | 2 | 3 | 2 | 2 | 2 |

Assumption 2 The total number of 'Married' and 'DeFacto' individuals is two times the number of family households that have two parents.

Assumption 3 The number of 'LoneParent' persons equals the number of single parent family households.

Assumption 4 The total number of 'GroupHhold' and 'LonePerson' individuals equals the number of residents in non-family households.

Assumption 5 The total number of 'Relative' and 'NonRelative' individuals equals to or is greater than the total number of males and females living in HF16 households.

Assumption 6 The total number of people for each type of family household is greater than or equal to the minimum number of people for that household type (calculated from Assumption 1 and table B24)

Assumption 7 If the number of households in any household category is 0 (table B24) and the corresponding number of males and/or females in table B25 is non-zero, values in table B25 will be set to 0.

Assumption 8 The total number of households that have children under 15 must be smaller than or equal to the number of total number of 'U15Child' individuals.

Assumption 9 The total number of households that have independent children older than 15 must be smaller than or equal to the total number of 'O15Child' individuals.

Assumption 10 The total number of households that have students must be smaller than or equal to the total number of ‘Student’ individuals.

Assumption 11 The number of non-family households that have 1 resident equals the number of ‘LonePerson’ individuals.

Assumption 12 The total number of ‘Relative’ and ‘NonRelative’ individuals from table B22a must be equal to or greater than 2 times the number of HF16 households.

Assumption 13 the total number of males and females from table B25a must be equal to the total number of males and females belonging to family households from table B22a.

Assumption 14 the total number of people in table belonging to family households (in table ‘Relationship in Household by Age by Sex’) must be equal to or greater than the minimum number of people required to live in the number of family households (in table ‘Family Composition’).

Assumption 15 the total number of ‘U15Child’, ‘Student’, ‘O15Child’, ‘LoneParent’, ‘Married’, ‘DeFacto’ individuals must be smaller than or equal to the number of residents allocated to households of types HF1 to HF15 (in table ‘Family Composition by Sex of Person in Family’).

After the corrections of mismatches in original ABS tables, table ‘Relationship in Household by Age by Sex’ was used to instantiate records in the pool of individual. Each record in this pool represents an individual person of the synthetic population and has four attributes, age, gender, household relationship and income. The age of an individual was randomly generated within the bounds of his/her age group. The three tables ‘Family Composition’, ‘Family Composition by Sex of Person in Family’, and ‘Household Composition by Number of Persons Usually Resident’ were used to instantiate records in household pool. Each record in this pool represents a household and has three attributes, number of males and number of females of the residents, and household type.

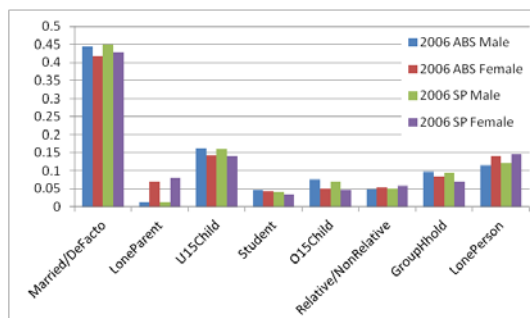


Figure 3. Percentage of males and females by household relationship of Randwick area in 2006

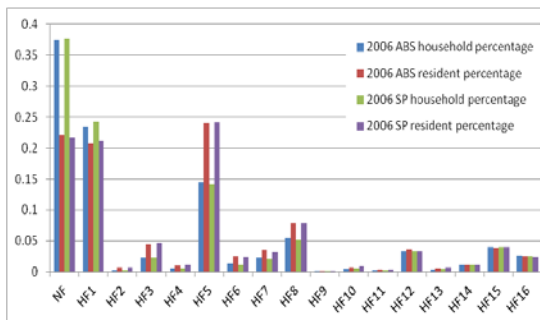


Figure 4. Percentage of 2006 household types in Randwick area

Sydney in year 2006 from ABS census data. In 2006 the area had approximately 106000 individuals living in around 47000 households who reside in private dwellings. Visitors were not included in the synthetic population because they were not considered as permanent residents in private dwellings in the study area.

Figure 3 compares the proportion of household relationships of male and female individuals in the synthetic population of Randwick area against the original ABS data. Figure 4 compares the proportion of the 17 household types in the synthetic population by number of households and number of residents against the original ABS data. Household types HF1 to HF16 are family households and are defined in Figures 2a and 2b. Group household members and lone persons live only in households of type NF (non-family households). These figures validate that the methodology presented in this paper can construct a realistic synthetic

2.1.2 Allocating individuals into households

The procedure to allocate individuals into households is as follow.

1. ‘Married’ and ‘DeFacto’ individuals are paired and randomly allocated to HF1 households and households having two parents (HF2 to HF8).
2. ‘LoneParent’ individuals are randomly allocated to households having single parent (HF9 to HF15)
3. ‘U15Child’ individuals, ‘Student’ individuals, and ‘O15Child’ individuals from the individual pool are consequently allocated to corresponding households, until the number of males and females of each household

in the household pool are reached.

4. Allocate ‘Relative’ individuals to HF16 households in the household pool.

5. Allocate the remaining ‘Relative’ and ‘NonRelative’ individuals to households HF1 to HF15 until the number of males and females of these households are reached.

6. Allocate ‘GroupHhold’ and ‘LonePerson’ individuals to NF households in the household pool.

2.1.3 Validating the initial synthetic population

The methodology presented here was implemented to construct a synthetic population for Randwick area in

population for agent based modelling purposes that matches very well with key statistics of the real population in the study area.

2.2 Synthetic population evolution algorithm

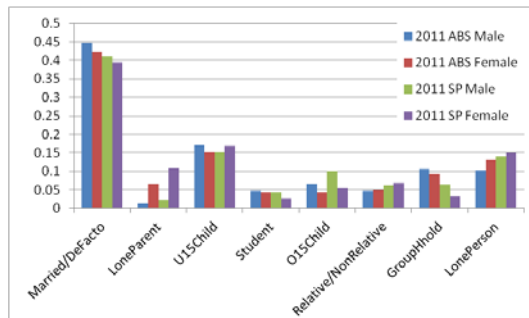


Figure 5. Percentage of males and females by household relationship of Randwick area in 2011

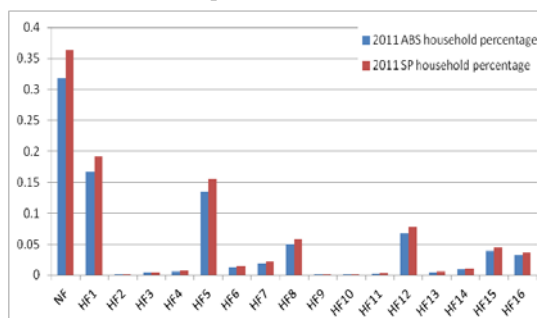


Figure 6. Percentage of household types (by number of households) in Randwick area in 2011

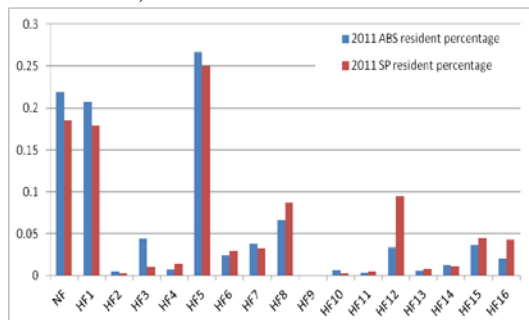


Figure 7. Percentage of household types (by residents) in Randwick area in 2011

household level. Simulation results show that in 2011 households of types NF, HF1, and HF5 maintain relatively higher proportions in the population compared to other household types. They also show a slight drop in the proportions of household types HF3 and HF12 compared to 2006, whereas those of other household types remain relatively unchanged. At individual level, the model successfully predicts that individuals in a relationship (married or de facto) occupy a higher proportion in the population compared to other individual types. The proportions of lone parents are relatively much lower, with significantly higher number of female lone parents compared to male lone parents.

There are however mismatches in the simulation results of evolution compared to 2011 ABS data. For example, the model failed to match the proportions of group household members and lone persons. It was also unable to predict the drop of non-family households in 2001 compared to 2006. The rates (for death, birth, marriage, and divorce) that drive the evolution through each of the 5 years in the simulation were for the whole of Australia in 2006 and thus when applied to the population in a small urban area (like Randwick) may produce discrepancies compared to the census data of the area. The algorithm for evolution will also need further refinements to better pick up the changes in household structure, for example adult children leaving their parent's house, and the formation of three (or more) generation households.

The attributes in Figure 5 are used to construct agents in an agent based model to simulate the evolution of the initial population. Each agent represents an individual, and individuals having the same 'Household ID' are grouped into a household. The evolution is simulated discretely to each individual on a yearly basis. In particular as part of the evolution, an individual ages, may die, may give birth to another individual, may divorce the individual he/she married to, and may marry another individual. The household relationship of each individual is re-evaluated during the course of this evolution (e.g. a 'Married' person undergoing a divorce should have his/her relationship in the household changed to 'LoneParent'). A divorce will always create a new household whereas a marriage may either form a new household (e.g. two 'O15Child' persons from two existing households marry each other), or remove a household (e.g. two 'LonePerson' individuals get married), or not change the total number of households (e.g. a 'O15Child' person marries to and moves in with a 'LoneParent' person). The household type attribute of a household is also re-evaluated during the course of the evolution to correctly reflect the composition of individuals in that household. Four sets of statistical data inform the processes in the evolution of the synthetic

population, which are death rates by age by gender, marriage rates by age by gender, divorcing rates by age by gender, and birth rates by age of females by number of babies born. The rates are available from ABS and are for the whole of Australia.

Statistics of the population presented in Section 2.1 after 5 years of evolution are shown in Figures 8 to 10, which match well with the statistics from ABS data for the Randwick LGA in year 2011. The agent based model, while simulating the evolution of the population discretely at individual level, was able to predict reasonably well the demographic distributions of the whole population in 2011, at individual level as well as

3. DISCUSSION AND CONCLUSIONS

This paper has presented an algorithm to construct a realistic synthetic population that uses only aggregated data of demographic distributions as inputs. It also presented an agent based model to simulate the natural evolutions (ageing, dying, marriage, divorce, reproducing) of this initial population. A synthetic population was constructed for Randwick area in Sydney in 2006 and evolved over 5 years. Comparisons of key statistics of the initial population against raw 2006 ABS data validated that the algorithm presented in this paper can construct a realistic synthetic population for agent based modelling purposes. Comparisons of key statistics of the population after evolution against 2011 raw ABS data proved that the agent based model developed in this paper while simulate the evolution of each individual discretely can predict reasonably well the demographics of the whole population at both household level and individual level.

The significance of the synthetic population developed in this work is in its ability to capture the relationship of individuals in a household and changes in structure of households as individuals undergo natural evolutions. Such a synthetic population is essential to the development of a simulation model to assist urban planning, particularly around capturing interactions of transport demand and relocation choice. This is because research has reported that transport related behaviours and decisions of individuals are highly influenced by the type and the composition the individuals have (e.g. co-travelling of individuals in the same household). Moreover, a synthetic population that ages and changes configuration will better reflect the future transport needs as well as dwelling needs.

REFERENCES

- Auld, J.A., Mohammadian, A., and Wies, K. (2009). Population synthesis with subregion-level control variable aggregation. *Journal of Transportation Engineering*. **135**: 632-639.
- Barthelemy, J., Toint, P.L., (2012). Synthetic population generation without a sample. *Transportation Science* (article in press).
- Beckman, R. J., Baggerly, K. A., and McKay, M. D. (1996). Creating synthetic baseline populations, *Transportation Research Part A: Policy and Practice*. 30 (6): 415-429.
- Guo, J. Y., and Bhat, C. R. (2007) Population synthesis for microsimulating travel behavior, *Transportation Research Record*, **2014** (12) 92-101.
- Frick, M., and Axhausen, K.W. (2004). Generating synthetic populations using IPF and Monte Carlo techniques: Some new results, paper presented at the 4th Swiss Transport Research Conference, Ascona, 2004.
- Harding, A., Lloyd, R., Bill, A., and King, A. (2004). Assessing poverty and inequality at a detailed regional level – New advances in spatial microsimulation, Research paper No. 2004/26, originally prepared for the UNU-WIDER Conference on Inequality, Poverty and Human Well-being, 30-31 May 2003, Helsinki
- Huang, Z., and Williamson, P. (2001). A comparison of synthetic reconstruction and combinatorial optimisation approaches to the creation of small-area microdata, working paper, Department of Geography, University of Liverpool.
- Melhuish, T., Blake, M., and Day, S. (2002). An evaluation of synthetic households populations for census collection districts created using spatial microsimulation techniques, prepared for the 26th Australia & New Zealand Regional Science Association International (ANZSRAI) Annual Conference, Gold Coast, Queensland, Australia, 2002.
- Muller, K., and Axhausen, K.W. (2011a) Population synthesis for microsimulation: State of the art, presented at the 90th Annual Meeting of the Transportation Research Board, Washington DC, 2011.
- Muller, K., and Axhausen, K.W. (2011b) Hierarchical IPF: Generating a synthetic population for Switzerland, presented at the 51st Congress of the European Regional Science Association, Barcelona, 2011.
- Ryan, J., Maoh, H., and Kanaroglou, P. (2009) Population synthesis: Comparing the major techniques using a small, complete population of firms, *Geographical Analysis*, **41** (2), 181-203.
- Voas, D., and Williamson, P. (2000). An evaluation of the combinatorial optimisation approach to the creation of synthetic microdata. *International Journal of Population Geography*. 6: 349-366.
- Williams, P. (2003). Using microsimulation to create synthetic small-area estimates from Australia's 2001 census", NATSEM working paper.
- Williamson, P., Birkin, M., Rees, P. (1998). The estimation of population microdata by using data from small area statistics and samples of anonymised records. *Environment and Planning A*. 30: 785-816.
- Wilson, A.G., and Pownall, C.E., (1976). A new representation of the urban system for modelling and for the study of micro-level interdependence, *Area*. 8: 246-254.
- Ye, X., Konduri, K., Pendyala, R.M., Sana, B., and Waddell, P. (2009). A methodology to match distributions of both household and person attributes in the generation of synthetic populations, presented at the 88th Annual Meeting of the Transportation Research Board, Washington DC, 2009.